



①9 BUNDESREPUBLIK
DEUTSCHLAND



DEUTSCHES
PATENTAMT

⑫ **Offenlegungsschrift**
⑩ **DE 197 51 536 A 1**

⑤1 Int. Cl.⁶:
G 10 L 7/08
G 01 H 3/00

②1 Aktenzeichen: 197 51 536.3
②2 Anmeldetag: 20. 11. 97
④3 Offenlegungstag: 27. 8. 98

DE 197 51 536 A 1

- | | |
|---|--|
| <p>③0 Unionspriorität:
5895/97 26. 02. 97 KR</p> <p>⑦1 Anmelder:
Samsung Electronics Co. Ltd., Suweon, Kyungki,
KR</p> <p>⑦4 Vertreter:
Grünecker, Kinkeldey, Stockmair & Schwanhäusser,
Anwaltssozietät, 80538 München</p> | <p>⑦2 Erfinder:
Ryu, Young-Moo, Kimcheon, Kyungsangbuk, KR</p> |
|---|--|

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

Prüfungsantrag gem. § 44 PatG ist gestellt

- ⑤4 Spracherkennungspegelsteuervorrichtung in einem Spracherkennungstelefongerät und zugehöriges Steuerverfahren
- ⑤7 Eine Vorrichtung für das Steuern des Spracherkennungspegels eines Spracherkennungstelefongerätes in Abhängigkeit vom Umgebungslärm umfaßt einen Ton-speicher für das Aufzeichnen von Sprachbefehlen, einen Lärmpegeldetektor für das Erkennen des Pegels des durch ein Mikrofon eingegebenen Umgebungslärmes, eine Steuerung für das Bestimmen des Spracherkennungspegels gemäß dem erkannten Lärmpegel, und einen Sprachbefehlserkennungsprozessor für das Verarbeiten des gesprochenen Sprachbefehls auf der Basis des bestimmten Spracherkennungspegels.

DE 197 51 536 A 1

HINTERGRUND DER ERFINDUNG

GEBIET DER ERFINDUNG

Die vorliegende Erfindung bezieht sich auf ein Spracherkennungstelefongerät und insbesondere auf eine Spracherkennungsvorrichtung und ein dafür verwendetes Steuerverfahren durch das Einstellen des Referenzspracherkennungspegels in Abhängigkeit vom Umgebungslärm, wenn der orale eingegebene (nachfolgend als "gesprochen" bezeichnet) Sprach-

befehl mit dem registrierten Sprachbefehl verglichen wird.

BESCHREIBUNG DES STANDES DER TECHNIK

Im allgemeinen ist das Spracherkennungstelefon ein Telefon, das Funktionen durchführen kann, wie Lauthören und ein Abspielen der automatischen Antwortnachricht in Erwiderung auf den gesprochenen Befehl des Benutzers.

Die Funktionsweise des konventionellen Spracherkennungstelefons wird nachfolgend beschrieben. Als erstes geht für den anfänglichen Registrierprozeß des Referenzsprachbefehls, wenn erkannt wird, daß die Funktionstaste und die Einstelltaste durch den Benutzer gedrückt werden, die Spracherkennungseinheit in die Sprachbefehlsaufzeichnungsbetriebsart, wobei, wenn Sprachbefehle durch das externe Mikrofon eingegeben werden und danach die Beendigungstaste gedrückt wird, die eingegebenen Sprachbefehle in digitale Daten umgewandelt werden, so daß sie aufgezeichnet/registriert werden. Wenn die Aufzeichnung solcher Sprachbefehle beendet ist, werden die aufgezeichneten Sprachbefehle abgespielt, um durch den Benutzer bestätigt zu werden. Beispielsweise stellt sich das Aufzeichnungsverfahren eines Sprachbefehls, der das Abspielen des automatischen Anrufbeantworters bewirkt, wie folgt dar: Drücken des Sprachbefehlsaufzeichnungsknopfes unter den Funktionstasten → Drücken des Einstellknopfes "*" → Drücken des beabsichtigten Funktionsknopfes (das ist der Abspielknopf) → Sprechen des Wortes "Abspielen" in das Mikrofon → Drücken des Beendigungsknopfes "#".

Um das Verfahren des Durchführens des gesprochenen Befehls zu beschreiben, wenn der gesprochene Befehl durch den Benutzer eingegeben wird, wird dieser mit dem aufgezeichneten/registrierten Sprachbefehl verglichen, wobei der gesprochene Befehl durchgeführt wird, wenn beide gleich sind. Die Identität des gesprochenen Befehls mit dem registrierten Sprachbefehl wird beurteilt durch Vergleichen von Sprachkennzeichen, basierend auf verschiedenen Frequenzen in Abhängigkeit von der Lautstärke und dem Ton.

Wenn beispielsweise der Benutzer "Abspielen" in das Mikrofon spricht, werden der gesprochene Befehl "Abspielen" und die registrierte Sprache "Abspielen" miteinander im Hinblick auf Sprachkennzeichen verglichen, und wenn es sich erweist, daß beide identisch sind, so wird die automatische Antwortvorrichtung gestartet, um aufgezeichnete Nachrichten abzuspielen, wohingegen wenn beide nicht identisch sind, der gesprochene Befehl "Abspielen" ignoriert wird.

Der Spracherkennungspegel, der als Referenz für das Beurteilen von Sprachkennzeichen verwendet wird, ist voreingestellt. Wenn beispielsweise der Spracherkennungspegel auf 70% eingestellt wird, so wird der gesprochene Befehl, wenn er zu 70% oder mehr mit dem registrierten Befehl identisch ist, akzeptiert. Wie oben erwähnt, muß im Falle eines höheren Spracherkennungspegels der gesprochene Befehl eine größere Identität mit dem registrierten Befehl haben, um akzeptiert zu werden, wohingegen im Falle eines niedrigeren Spracherkennungspegels der gesprochene Befehl trotz seiner niedrigeren Identität mit dem registrierten Sprachbefehl akzeptiert werden kann.

Das heißt, wenn der Spracherkennungspegel höher ist, so kann der gesprochene Befehl korrekt ausgeführt werden, aber die Erkennungsrate ist geringer, und im Gegensatz dazu ist, wenn der Spracherkennungspegel niedriger ist, die Erkennungsrate höher, aber es werden Fehlfunktionen verursacht.

Das konventionelle Spracherkennungstelefon, wie es oben erwähnt wurde, hat den Nachteil, daß die Spracherkennungsrate eines gesprochenen Befehls niedriger wird durch die Lärmischung im Falle eines niedrigeren Signal-Rausch-(S/N)-Verhältnisses, wenn der Benutzer den anfänglichen Sprachbefehl als Referenzspracherkennungspegel registriert. Weiterhin wird im Falle eines niedrigeren S/N-Verhältnisses des tatsächlich gesprochenen Befehls die Spracherkennungsrate niedrig, obwohl der Referenzspracherkennungspegel anfänglich mit einem höheren S/N-Verhältnis aufgezeichnet wurde.

ZUSAMMENFASSUNG DER ERFINDUNG

Somit besteht eine Aufgabe der vorliegenden Erfindung darin, eine Spracherkennungsvorrichtung zu liefern, die den Spracherkennungspegel in Abhängigkeit vom Lärmpegel, der mit einem gesprochenen Befehl gemischt ist, einstellen kann, um die Spracherkennungsrate zu verbessern.

Eine andere Aufgabe der vorliegenden Erfindung besteht darin, ein Verfahren für das Steuern des Spracherkennungspegels in Abhängigkeit vom Lärmpegel, der mit einem gesprochenen Befehl gemischt ist, zu steuern, um die Spracherkennungsrate zu verbessern.

Gemäß der vorliegenden Erfindung umfaßt eine Vorrichtung für das Steuern des Spracherkennungspegels eines Spracherkennungstelefongerätes in Abhängigkeit vom Umgebungsgerauschen einen Tonspeicher für das Aufzeichnen der Sprachbefehle, einen Lärmpegeldetektor für das Erkennen des Pegels des Umgebungslärms, der durch das Mikrofon eingegeben wird, eine Steuerung für das Bestimmen des Spracherkennungspegels gemäß dem erkannten Lärmpegel, und einen Sprachbefehls-erkennungsspektor für das Verarbeiten einen gesprochenen Sprachbefehls auf der Basis des bestimmten Spracherkennungspegels.

Fig. 1 ist ein Blockdiagramm, das ein Spracherkennungstelefongerät gemäß einer Ausführungsform der vorliegenden Erfindung zeigt; und

Fig. 2 ist ein Flußdiagramm, das das Verfahren des Steuerns des Spracherkennungspegels gemäß einer Ausführungsform der vorliegenden Erfindung zeigt.

DETAILLIERTE BESCHREIBUNG DER BEVORZUGTEN AUSFÜHRUNGSFORM

Die Konfiguration eines Spracherkennungstelefongerätes gemäß einer Ausführungsform der vorliegenden Erfindung wird nachfolgend unter Bezug auf **Fig. 1** beschrieben. Die Steuerung **10** steuert den gesamten Betrieb des Spracherkennungstelefongerätes mittels des im Speicher **34** gespeicherten Steuerprogramms. Der Sprechkreis **14** bildet eine Schnittstelle verschiedener Tonsignale, Sprachsignale, etc. unter der Steuerung der Steuerung **10**. Das Gabelrelais H/S bildet oder unterbricht mechanisch den Sprachkanal zwischen dem Sprechkreis **14** und der Telefonleitung, wenn der Handapparat abgehoben wird. Das Sprechrelais bildet oder unterbricht den Sprachkanal zwischen dem Sprechkreis **14** und der Telefonleitung unter der Steuerung der Steuerung **10**. Das Mikrofon **16** wandelt das Sprachsignal in ein elektrisches Signal, und der Lautsprecher **18** wandelt das elektrische Signal in ein Tonsignal, um einen hörbaren Ton zu erzeugen. Der Lärmsignaldetektor **34** ist mit der Telefonleitung verbunden, die zum Gabelrelais H/S führt, um somit das Lärmsignal zu erkennen, das durch die Telefonleitung empfangen wird, um es weiter an die Steuerung **10** zu geben.

Die ersten und zweiten CODECs **20, 22** wandeln das analoge (Ton-) Signal in ein digitales Signal und umgekehrt. Der Sprachbefehlserkennungsprozessor **24** verarbeitet den digitalisierten Sprachbefehl, der von den ersten und zweiten CODECs **20, 22** empfangen wird, und zeichnet das sich ergebende Signal im Tonspeicher **26** auf, oder er erkennt den Eingabesprachbefehl auf der Basis des Spracherkennungspegels, der von der Steuerung **10** geliefert wird. Wenn der Eingabesprachbefehl mit dem registrierten Sprachbefehl, der im Tonspeicher **26** gespeichert ist, identifiziert wird, so überführt der obige Sprachbefehlserkennungsprozessor **24** Daten, die die Ausführung des Eingabesprachbefehls bewirken, an die Steuerung **10**.

Der Lärmpegeldetektor **28** besteht aus einem DC-Gleichrichter (Gleichstromgleichrichter) **30** und einem A/D-Wandler **32**, um somit das Pegelsignal als numerische Daten gemäß dem Tonsignalpegel, der vom externen Mikrofon **16** eingegeben wird, zu erzeugen. Der DC-Gleichrichter **30** wandelt das analoge Tonsignal, das vom Mikrofon **16** angelegt wird, in ein DC-Signal (Gleichstromsignal). Der A/D-Wandler **32** wandelt den Signalpegel, basierend auf der Größe des DC-Signals, das durch den DC-Gleichrichter **30** angelegt wird, in ein digitales Signal in Form numerischer Daten, um so das Pegelsignal auszugeben. Gewöhnlicherweise wird der A/D-Wandler, der in der Steuerung vorgesehen ist, verwendet, oder es kann ein getrennter A/D-Wandler zusätzlich vorgesehen werden. Der Speicher **36** speichert Steuerprogramme und Daten, die vom Betrieb des Telefongerätes herrühren. Das Tasteneingabefeld **38** erzeugt Tastenbefehle und Tastendaten für das Steuern des Betriebs des Telefongerätes. Das Anzeigefenster **40** zeigt die aktuellen Betriebszustände des Telefons unter der Steuerung der Steuerung **10** an.

Nachfolgend wird der Betrieb eines Spracherkennungstelefongerätes gemäß einer Ausführungsform der vorliegenden Erfindung im Detail unter Bezug auf **Fig. 1** erläutert. Als erstes geht für das anfängliche Registrierungsverfahren des Referenzsprachbefehls, wenn erkannt wurde, daß die Funktionstaste und die Einstelltaste durch den Benutzer in der Wartestellung gedrückt wurden, der Sprachbefehlserkennungsprozessor **24** in die Sprachbefehlsaufzeichnungsbetriebsart über, wobei der durch den Benutzer eingegebene Referenzsprachbefehl zum ersten CODEC **20** geliefert wird, der den obigen analogen Sprachbefehl in einen digitalen Sprachbefehl umwandelt, um daraufhin den digitalisierten Sprachbefehl an den Sprachbefehlserkennungsprozessor **24** zu liefern. Und wenn der Benutzer den Beendigungsknopf auf dem Tasteneingabefeld **38** drückt, speichert der Sprachbefehlserkennungsprozessor **24** einen solchen digitalisierten Sprachbefehl im Tonspeicher **26** unter der Steuerung der Steuerung **10**.

Unter Bezug auf **Fig. 2** wird der Betrieb des Spracherkennungstelefongerätes in Erwiderung auf die Anforderung für das Ausführen des durch den Benutzer eingegebenen Sprachbefehls nachfolgend beschrieben. Als erstes wird das Verfahren des Erkennens des Lärmpegels in den Schritten **214-220** beschrieben. Wenn in Schritt **214** ein spezifiziertes Zeitintervall vergangen ist, so geht die Steuerung **10** zu Schritt **216**. Die Prüfung des Ablaufs eines spezifizierten Zeitintervalls (gewöhnlicherweise einige Millisekunden) dient dazu, das Lärmpegelerkennungsverfahren periodisch in solchen Zeitintervallen durchzuführen. In Schritt **216** erkennt die Steuerung **10** den Signalpegel basierend auf der Größe des Lärms, der vom Mikrofon **16** eingegeben wird.

Wie beim Lärmpegelerkennungsverfahren wird das vom Mikrofon **16** eingegebene Lärmsignal an den DC-Gleichrichter **30** gegeben, wobei die Direktstromkomponente eliminiert wird, und das sich ergebende Lärmsignal ohne Gleichstromkomponente an den A/D-Wandler **32** geliefert wird, wodurch es in binäre Zifferndaten umgewandelt wird, die aus einer spezifizierten Zahl von Bits bestehen, die Lärmpegel entsprechen, um dann weiter zur Steuerung **10** übertragen zu werden. Die Zahl der Bits, die die obigen Lärmpegel darstellt, wird durch die Zahl der Stufen der Spracherkennungspegel bestimmt. Beispielsweise können, wenn die Spracherkennungspegel in vier Stufen klassifiziert sind, die Sprachpegel durch binäre Zifferndaten von zwei Bits dargestellt werden. Wie oben beschrieben wurde, geht, wenn der Lärmpegel in Schritt **216** erkannt wird, die Steuerung **10** zu Schritt **218** weiter, um ein solches Lärmsignal im Speicher **36** oder einem internen Hilfspuffer zu speichern.

Nach dem Speichern eines solchen erkannten Lärmpegelsignals, geht die Steuerung **10** zu Schritt **220** weiter, um dem Sprachbefehlserkennungsprozessor **24** den relevanten Spracherkennungspegel, der dem so erkannten Lärmpegel entspricht, mittels einer Lärmpegelstufentabelle, die im Speicher **36** aufgezeichnet ist, mitzuteilen. Wenn die Steuerung **10** beispielsweise binäre Zifferndaten "01" vom Lärmpegeldetektor **28** empfängt, so liefert sie den Spracherkennungspegel der zweiten Stufe, wie das in der folgenden Tabelle 1 aufgelistet ist, an den Sprachbefehlserkennungsprozessor **24**. Ein Beispiel der Spracherkennungspegel gemäß den Lärmpegeln ist in der folgenden Tabelle 1 aufgelistet.

5	Lärm- pegel	Spracherken- nungspegel	Lärmzustand
	00	1. Stufe	Zustand wie anfänglich registriert
10	01	2. Stufe	Gespräch möglich
	10	3. Stufe	Gespräch ungünstig
	11	4. Stufe	Spracherkennung unmöglich

15 Als nächstes wird das Verfahren des Ausführens des Sprachbefehls in den Schritten **212-222** beschrieben. In Schritt **212** prüft die Steuerung **10**, ob der Spracherkennungsbefehl vom Sprachbefehlserkennungsprozessor **24** empfangen wird.

20 Um das Verfahren der Ausgabe des Sprachbefehls vom Sprachbefehlserkennungsprozessor **24** zu beschreiben, wird der vom Mikrofon **16** eingegebene Sprachbefehl zunächst an den CODEC **20** angelegt, um in ein digitales Signal umgewandelt zu werden. Dann wird der vom ersten CODEC **20** ausgegebene digitalisierte Sprachbefehl weiter an den Sprachbefehlserkennungsprozessor **24** angelegt, der dann einen solchen digitalisierten Sprachbefehl mit dem Spracherkennungspegel, der von der Steuerung **10** empfangen wird, nach dem Durchführen von Schritt **220** vergleicht. Da ein solches Sprachbefehlserkennungsverfahren durch Vergleichen des eingegebenen (gesprochenen) Sprachbefehls mit dem registrierten Sprachbefehl eine konventionelle Technik für die Spracherkennungstelefongeräte darstellt, wird hier keine detaillierte Beschreibung gegeben. Zusammenfassend kann man sagen, wenn der Spracherkennungspegel der zweiten Stufe, auf den durch die Steuerung **10** Bezug genommen wird, beispielsweise auf 80% eingestellt wird, so prüft der Sprachbefehlserkennungsprozessor **24**, ob sich die Sprachkennzeichen des gesprochenen und des gespeicherten Sprachbefehls zu 80% oder mehr ähneln. Wenn ermittelt wird, daß die gesprochenen und registrierten Sprachbefehle identisch sind, so liefert der Sprachbefehlserkennungsprozessor **24** den Spracherkennungsbefehl an die Steuerung **10**.

30 Daraufhin geht die Steuerung **10**, wenn sie den Spracherkennungsbefehl in Schritt **212** empfängt, zu Schritt **222**, um einen solchen Spracherkennungsbefehl nach der auszuführenden Funktion zu analysieren, und nach Beendigung der Analyse geht die Steuerung **10** zu Schritt **224** weiter, um die als Ergebnis der Analyse angeforderte Funktion auszuführen. Es erfolgt keine detaillierte Beschreibung des Verfahrens der Schritte **222** und **224**, da sie auf einer konventionellen Technik beruhen.

35 Wie oben beschrieben wurde, hat die vorliegende Erfindung den Vorteil, daß der Spracherkennungspegel unter Berücksichtigung des Umgebungslärmpegels eingestellt werden kann, um somit den Effekt des Umgebungslärms zu minimieren und somit die Spracherkennungsrate zu verbessern.

40 Patentansprüche

1. Vorrichtung zur Steuerung des Spracherkennungspegels eines Spracherkennungstelefongerätes in Abhängigkeit vom Umgebungslärm, umfassend:
 einen Tonspeicher für das Aufzeichnen von Sprachbefehlen;
 45 einen Lärmpegeldetektor für das Erkennen des Pegels des Umgebungslärmes, der durch das Mikrofon eingegeben wird;
 eine Steuerung für das Bestimmen des Spracherkennungspegels gemäß dem erkannten Lärmpegel; und
 einen Sprachbefehlserkennungsprozessor für das Verarbeiten des gesprochenen Sprachbefehls auf der Basis des bestimmten Spracherkennungspegels.
2. Vorrichtung zur Steuerung des Spracherkennungspegels in Abhängigkeit vom Umgebungslärm nach Anspruch 1, wobei der Lärmpegeldetektor folgendes umfaßt:
 einen DC-Gleichrichter (Gleichstromgleichrichter) für das Eliminieren der Gleichstromkomponente aus dem Umgebungslärm; und
 50 einen Analog/Digital-Wandler für das Umwandeln des Umgebungslärms ohne die Gleichstromkomponente in binäre Zifferndaten, die dem Lärmpegel entsprechen.
3. Verfahren in einem Spracherkennungstelefongerät, das einen Lärmpegeldetektor für das Erkennen des Pegels des Umgebungslärms und einen Sprachbefehlserkennungsprozessor für das Verarbeiten eines gesprochenen Sprachbefehls auf der Basis des Spracherkennungspegels umfaßt, zur Steuerung des Spracherkennungspegels in Abhängigkeit vom Umgebungslärm, wobei das Verfahren folgende Schritte umfaßt:
 60 periodisches Erkennen des Lärmpegels durch den Lärmpegeldetektor in spezifizierten Zeitintervallen; und
 Versorgen des Sprachbefehlserkennungsprozessors mit dem Spracherkennungspegel, der dem Lärmpegel entspricht.
4. Verfahren in einem Spracherkennungstelefongerät, das einen Lärmpegeldetektor für das Erkennen des Pegels des Umgebungslärmes und einen Sprachbefehlserkennungsprozessor für das Verarbeiten eines gesprochenen Sprachbefehls auf der Basis des Spracherkennungspegels umfaßt, für das Steuern des Spracherkennungspegels in Abhängigkeit vom Umgebungslärm, wobei das Verfahren folgende Schritte umfaßt:
 65 periodisches Erkennen des Lärmpegels zu spezifizierten Zeitintervallen mittels des Lärmpegeldetektors;
 vorübergehendes Speichern des erkannten Lärmpegels; und

Versorgen des Sprachbefehlserkennungsprozessors mit dem Spracherkennungspegel, der dem erkannten Lärmpegel entspricht.

Hierzu 2 Seite(n) Zeichnungen

5

10

15

20

25

30

35

40

45

50

55

60

65

- Leerseite -

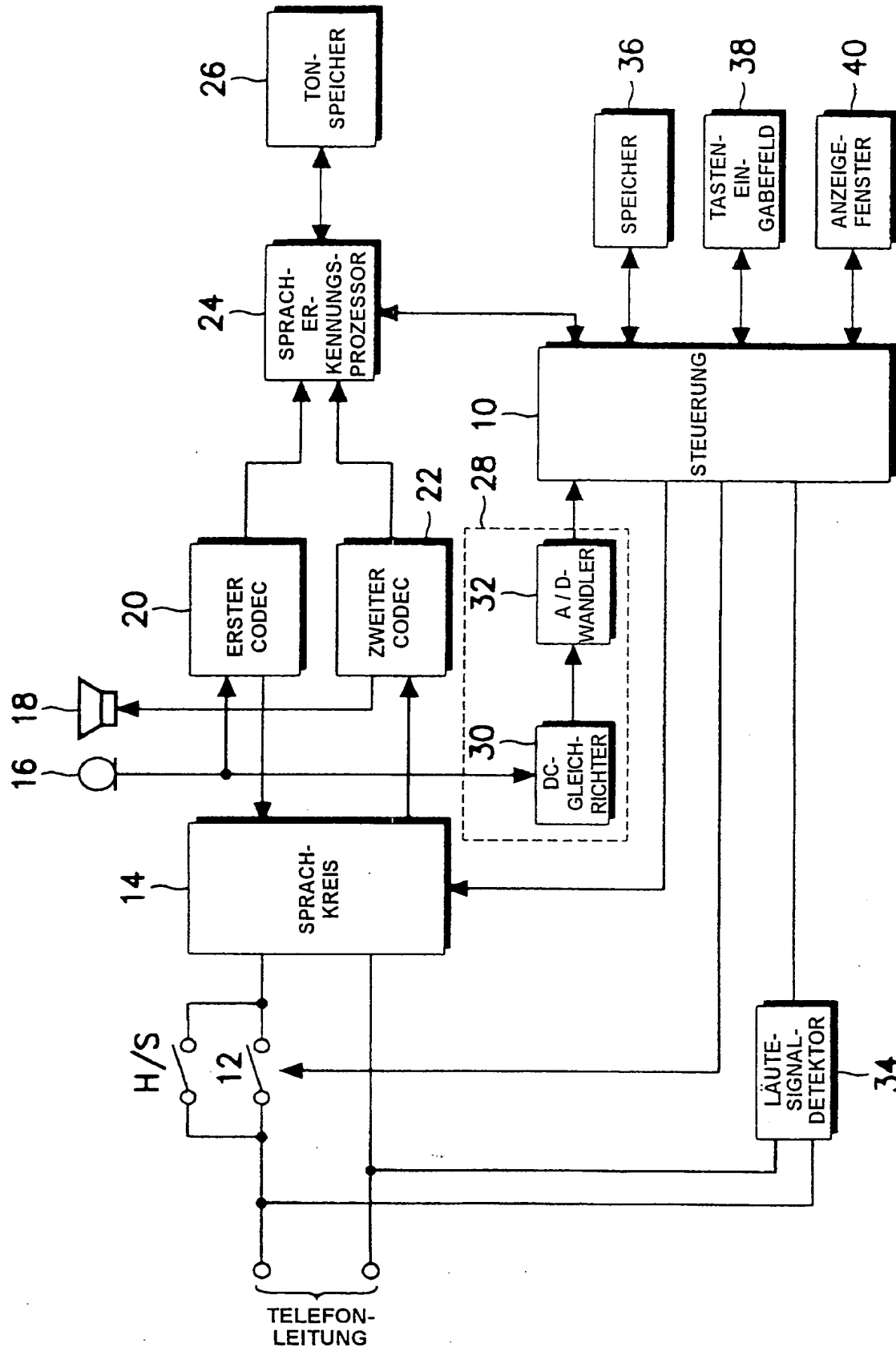


FIG. 1

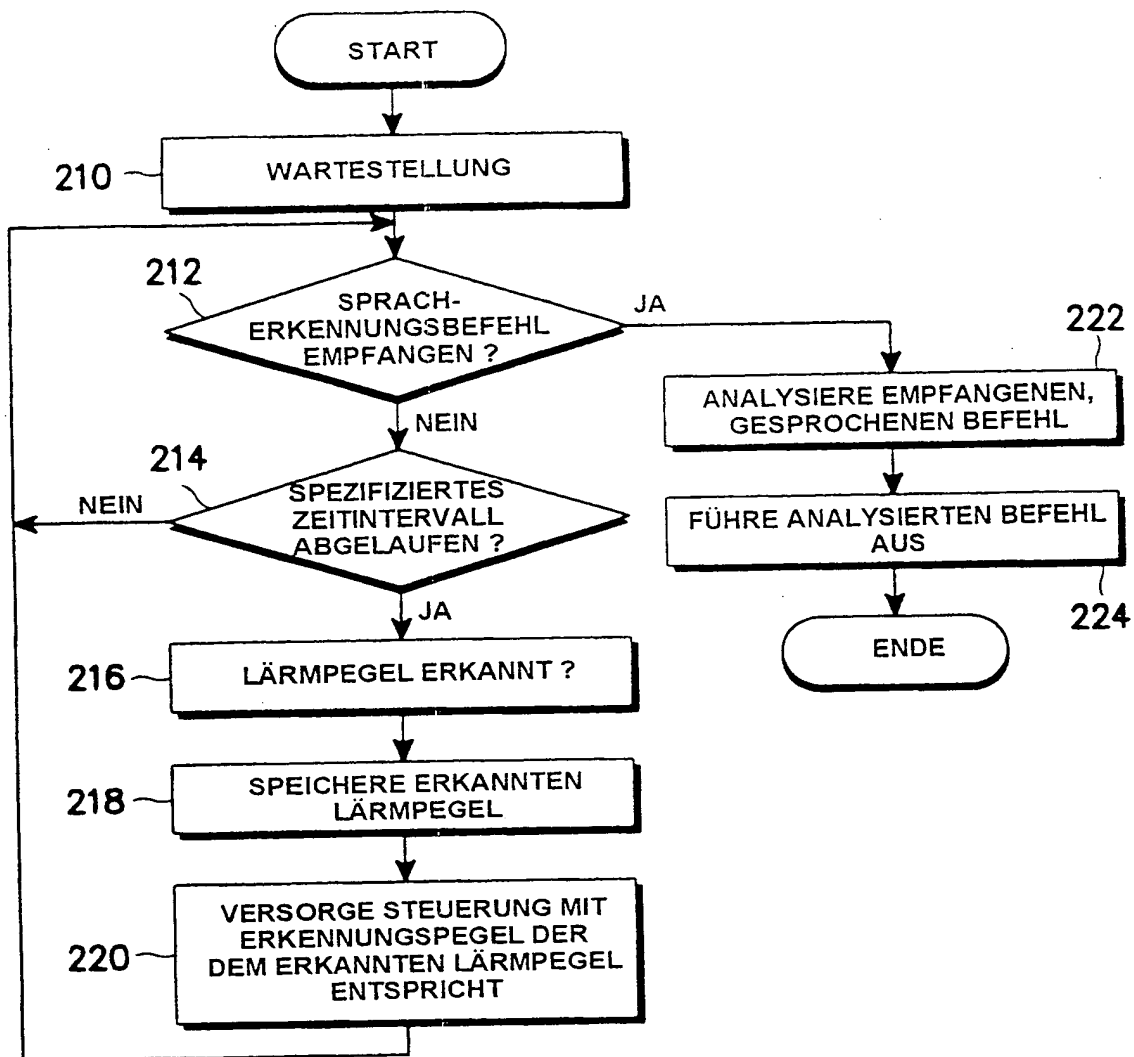


FIG. 2